

国立国語研究所学術情報リポジトリ

従属節の意味分類基準策定について：鳥バンク基準互換再構築の検討

著者	松本 理美
雑誌名	言語資源活用ワークショップ発表論文集
巻	2
ページ	40-51
発行年	2017
URL	http://doi.org/10.15084/00001504

従属節の意味分類基準策定について ー鳥バンク基準互換再構築の検討ー

松本 理美 (立命館大学大学院文学研究科)

Defining Annotation Standard of Clause Boundaries and Its Semantic Classes

Study on Re-consideration of the Clause Classes in Tori-bank

Satomi Matsumoto (Graduate school of Letters, Ritsumeikan University)

要旨

本発表は、「鳥バンク」基準互換 (池原, 2009) (以下「鳥バンク」) の日本語従属節の意味分類¹基準の策定について再検討を行う。「鳥バンク」は日英対訳コーパス中の文型パターン
の分析・分類を重ねグループ化された意味分類体系である。しかし、日本語ー英語翻訳を意
識した分類体系であるため、そのまま日本語分析のための従属節意味分類に転用すること
には若干の問題がある。『現代日本語書き言葉均衡コーパス』に対する「鳥バンク」基準の
従属節アノテーション作業において、二人のアノテーターによる意味分類タグ付け作業の
齟齬の分析を通し、日本語分析のための基準の再構築を試みる。具体的には、意味・機能の
階層の修正 (分類階層に混在している意味と機能の整理)、分類項目 (日本語学的視点で節
末形式や文例の再分類) についての議論を行う。アノテーション作業における問題点の整理
から得た知見を基に、「鳥バンク」基準の再構築による安定した日本語の従属節分類基準策
定について検討する。

1. はじめに

本発表は、日本語ー英語翻訳を意識した「鳥バンク」由来の従属節意味分類アノテーシ
ョン基準 (後述) を日本語に特化した基準とするために、日本語学の視点から修正を提案する
ことを目的とする。『現代日本語書き言葉均衡コーパス』 (以下 BCCWJ) に対する「鳥バン
ク」基準の意味分類タグ付け作業で生じた作業者間の齟齬データを再分析し、タグ付与齟齬
の要因と考えられる「鳥バンク」意味分類体系の問題を明らかにする。そして、「鳥バンク」
を日本語仕様の意味分類体系に再構築するための修正について議論する。

昨今のコーパス開発の急進により、機械学習に基づく言語解析器の開発、アノテーション
の自動化が進んでいるが、従属節の意味分類については、人手に頼るところが大きい。日本
語の従属節の意味については日本語文法の視点から数多くの研究者による膨大な理論的研
究があるが、日本語の従属節の意味分類の自動化について丸山 (2016) に代表されるような
日本語学者と工学者の協働の歴史はまだ浅い。

大量の文型パターンをグループ化することにより作成された「鳥バンク」節間意味分類体
系は 222 種類の項目から成り、600 余りの「節間キーワード」と日本語の文例が各項目に分
類された非常に価値の高いものである。「鳥バンク」を手掛かりとした BCCWJ に対する意
味分類のタグ付け作業で生じた作業者間の意味分類の齟齬は、最終的に 10%程度であり決

¹ 池原 (2009) は「節の意味分類」について、「節は単一の事象を表現したもので、その内容は単文に相
当するため、節の意味分類は単文の意味分類に従う」 (p.265) とし、節が表現している事象について
の意味分類を指している。本発表で着目している従属節の意味分類を「節間の意味分類」としているが、本
発表では「従属節の意味分類」とする。

して多いものではない。しかしながら、「鳥バンク」は日英・英日自動翻訳にも対応するよう工学的視点から開発されており、そのまま日本語従属節の意味分類に適用すると問題が生じることが明らかになった。意味分類の1段階で生じた齟齬の9割が副詞節ー並列節で生じ、その他の齟齬は補足節の引用節に関連した分類で生じている。2段階以下の意味分類では、名詞修飾節内の内容節ー縮約形修飾節、副詞節内の手段ー因果関係、時ー条件、そして因果関係の下位分類の原因ー結果などで生じている。「鳥バンク」の従属節意味分類体系は日本語の分析に有用であると考えられるが、一部に意味分類体系が要因で生じたと考えられる齟齬があることも事実である。

本発表では、「鳥バンク」を手掛かりとしたBCCWJに対する意味分類のタグ付け作業で齟齬が生じた実例を示しながら、従属節の機能・意味による階層、分類項目、節末形式や用例（「節間キーワード」や日本語の文例）を見直し、「鳥バンク」を日本語仕様に再構築するための修正について検討する。

2. 「鳥バンク」基準の従属節の意味分類アノテーションに見られる問題

まずBCCWJの一部に対する「鳥バンク」基準の従属節の意味分類アノテーション作業について、松本・浅原・有田（2017）で整理した問題を述べ、次に池原（2009）に基づき「鳥バンク」の目的、分類体系の構築過程について概観する。そして、「鳥バンク」基準と意味分類アノテーション作業で生じた作業間齟齬の関連について事例を挙げて説明する。

2. 1 従属節の意味分類アノテーション作業の齟齬分析

松本・浅原・有田（2017）では、BCCWJの新聞記事コアデータの一部を対象に「鳥バンク」を意味分類基準とした従属節の意味分類アノテーション作業についての問題点を整理した。1次作業で生じた作業間齟齬を分析し、従属節の意味分類アノテーションの基準策定の必要性について以下のように述べた。

- i) タグ付け齟齬の約57%が節認定において生じ、その77%が表1に示す通り名詞修飾節と副詞節において生じた。

表 1 節認定の齟齬の頻度²

補足節	名詞修飾節	副詞節	並列節
102 (16%)	275 (44%)	207 (33%)	47 (7%)

節認定の問題に関しては、①何を節として認定するか ②どこからどこまでを節として認定するか という両面から論じる必要がある。述語の性質、意味から節を認定することは極めて難しく、厳密に節認定を行おうとすれば、認定はますます複雑化することが予想される。節認定に関して、文法上の議論の余地がある事例においても、広めの許容範囲を認め、節情報タグを付与する方向での基準策定の可能性を探ることが求められる。

² 節認定の齟齬とは、作業員2名のうち、1名が従属節と認定してアノテーションを行ったが、もう1名がアノテーションを行わなかったものをいう。

- ii) 従属節の意味分類³の齟齬については、機能、意味に分けて分析した。作業者により認定した従属節の種類が異なるという従属節の機能における分類齟齬の頻度は表 2 の通りである。

表 2 従属節の機能分類における作業者間齟齬の頻度 (1 次作業)

A \ B	補足節	名詞修飾節	副詞節	並列節
補足節	8	4	7	8
名詞修飾節	11	179	4	0
副詞節	12	6	85	125
並列節	1	0	26	2

従属節の機能における分類の齟齬は並列節－副詞節で最も多く、従属節機能分類の齟齬の 74%を占めている。2 段階以下の意味分類の齟齬頻度は表 2 の着色した部分であるが、最多となる名詞修飾節においては、補足語修飾節間（限定的－非限定的）と内容節－補足語修飾節－縮約形修飾節間という下位分類での齟齬が 75%を占める。

2. 2 「鳥バンク」について

池原 (2007, 2009) によると、「鳥バンク」は日本語の重文と複文を対象とした、大規模な「意味類型パターン辞書」(22.7 万件) を収録したデータベースである。「重文・複文の持つ統語的な構造と意味の関係に着目」(池原, 2009:254) し、粗いレベルでの従属節の意味分類の半自動化の可能性を追求している。意味的等価変換方式⁴の実現を目指したもので、これに必要となる重文・複文の意味が分類体系化されている。ここでは日本語表現の意味が英語表現で定義されている文型パターンが分類されている。池原 (2009) は「日本語表現の意味分類の粒度として、英語表現との意味的対応が取れる範囲以上の物は不要である」(p.257) としている。つまり、「鳥バンク」の意味分類体系は、英語への翻訳を意識したもの⁵であることを確認しておく必要がある。

主節と従属節の意味的关系を分類した「節間の意味分類」について、その構築の過程を簡潔に述べる。構築過程は大きく 3 段階に分けられる。まず 1 つ目の段階は、「統語的特徴を考慮した大分類」である。益岡・田窪 (1992)、益岡 (1997) を参考に、従属節を「補足節（補語相当節）」「名詞修飾節（連体節）」「副詞節（副詞的連用節）」「並列節（並列型連用節）」に 4 分類した後、2 段階でそれを「統語的色彩の強いレベル」でさらに 27 分類している。次に 3 段階を「節間キーワードを考慮した中分類」として⁶、「約 1000 件の文型パターンを対象に従属節と主節とを連結するキーワードとその意味に着目した用例分析を行い、分類を詳細化」(p.259) して 37 の下位分類を行っている。最後は 4 段階の「用例分類過程での小

³ 「鳥バンク」では、従属節の機能と意味を厳密に分類していない。本発表では、鳥バンクの 1 段階の分類に従属節の機能における分類と解釈した。

⁴ 「概念の持つ構造と表現形式の関係に着目したもので、原言語の文型パターンと目的言語の文型パターンそれぞれを意味類型化（意味的に分類）し、両言語間で同一の意味を持つ文型パターングループ間で写像を行う方法である。」(池原, 2009:253)

⁵ 同じ言語内での変換（言い換え）も含んでいる。

⁶ 池原 (2009) による大中小の分類について、本文中の説明と表 8.5 (p.264) に齟齬があると思われるので、本文中の説明を優先的に解釈した。

分類」である。3段階までで「得られた分類体系を12万件の文型パターン辞書（単語レベル）に適用し、文型パターンを実際にグループ化した。」(p.265) 分類項目の新設や細分化により154の下位分類を行い、「最終的に222種類からなる意味分類体系を作成」(p.265) している。さらに、「節間キーワード」と対応させた600余りの日本語の文例を提示している。

2.3 「鳥バンク」と意味分類アノテーション作業の齟齬

本発表においては、前節で述べた松本・浅原・有田 (2017) の指摘を踏まえ、特に意味分類作業において齟齬が生じたものに着目し、アノテーション作業の基準とした「鳥バンク」意味分類体系に起因した意味分類の齟齬について述べる。1次作業で2名の作業者が付与したタグの齟齬と、2次作業において付与したタグとの齟齬を再検討した。明らかなタグの誤入力、誤分類と判断できるものを除き、意味分類の揺れとして生じた齟齬のみを再抽出した。再抽出された齟齬の頻度は表3の通りであった。

表3 再検討後の従属節機能分類における齟齬の頻度

A \ B	補足節	名詞修飾節	副詞節	並列節
補足節	0	0	0	0
名詞修飾節	6	10	0	0
副詞節	5	0	42	79
並列節	0	0	9	1

再抽出した152例の齟齬のうち、最も大きな分類である1段階の従属節の機能分類で生じた齟齬は99例、2段階以下の分類において齟齬が生じたのは53例であった。

以下に、「鳥バンク」意味分類体系と意味分類タグ付け作業の齟齬の関連について事例を挙げながら詳述する。

2.3.1 従属節の機能分類における齟齬

まず、従属節の機能分類における齟齬について述べる。従属節の機能分類の齟齬とは、タグ付け作業において作業者Aが補足節、作業者Bが副詞節とタグ付けしたような従属節間の齟齬を指し、従属節の意味分類アノテーション基準において、1段階（大分類）での齟齬である。頻度は表3の着色していない部分の通りであり、再抽出により1次作業時の齟齬の5割に留まったが、その9割が副詞節－並列節に見られることは注目に値する。以下に具体例を挙げて論点を明らかにする。

なお、事例の前行に作業者が「鳥バンク」を基準に付与したタグを、作業者A、作業者Bの順で《1段階（2段階：3段階）－1段階（2段階：3段階）》と示した。また、事例は付与されたタグに齟齬のあった従属節を含む文を示し、アノテーションした語に下線を付したが、長すぎる文については文脈判断に支障がない範囲で前後を省略した。文の後には作業ファイル名⁷を記した。

⁷ 作業ファイルは <https://github.com/masayu-a/BCCWJ-ANNOTATION-ORDER> 参照

2017年9月5日-6日

- (4) 昨年十月から二カ月間、アジア六カ国・地域を回り、中国で出会った無名の日本人ミュージシャンの話に驚いた。[PN4g_00002, file00029]
- (5) 「三遊亭王楽の天狗道場・銀座編～円楽と東西の若手達」と題して、三遊亭円楽と王楽のほか、桂つく枝、柳家三三が出演する。[PN4g_00002, file00030]
- (6) 研究会は PTA 広報紙や学校新聞などを作る際の問い合わせにも応じ、情報を提供している。[PN5a_00002, file00031]

前述したとおり、従属節の機能分類における齟齬では、副詞節－並列節の齟齬が最も多く 9 割を占めている。作業者 B は(3)～(6)を並列節（順接的並列：総記の並列）としているが、作業者 A は(3)を副詞節（手段）、(4)を副詞節（手段 or 付帯状況）、(5)を副詞節（付帯状況）、(6)を副詞節（手段 or 付帯状況）としている。並列節と副詞節の齟齬の場合、副詞節タグの多くは、手段・付帯状況・因果関係（原因）・因果関係（結果）のいずれを選択するかで揺れを示し、他のタグを欄外に注記しているものも確認された。

本発表では副詞節－並列節の齟齬について、並列節と副詞節を連用節の下位に分類するという丸山 (2016) の立場をとることを提案するため（後述）、ここでは事例を見るに留める。

なお、1 段階で連用節、下位の 2 段階で副詞節と並列節に分類すると、1 段階では副詞節と並列節の分類が一旦解消されるため、表 3 の右下の 4 つのセルが統合される。従属節の機能を補足節・名詞修飾節・連用節の 3 分類とした場合、従属節の機能により分類した場合の齟齬は事例(1)(2)で示した名詞修飾節－補足節、副詞節－補足節の 11 例ということになる。この 11 例は、いずれも引用節が関与しているため、引用節の扱いを修正することで、1 段階の意味分類の齟齬は解消する。

2. 3. 2 意味分類における齟齬

次に 1 段階の機能分類において齟齬がなく、下位の 2 段階で齟齬が生じた名詞修飾節、副詞節、並列節について従属節ごとに事例を挙げて述べる。

A. 名詞修飾節における 2 段階以下の齟齬

名詞修飾節の下位分類での齟齬は、2 段階以下での齟齬の 2 割に満たない。その齟齬の 10 例中 8 例は、2 段階の内容節－縮約形修飾節で生じている。「鳥バンク」では、内容節を「被修飾名詞の内容を表す節。被修飾名詞と修飾節が同格の関係にあるタイプ。」、縮約形修飾節を「被修飾名詞と修飾節が格関係にも、同格関係にもなく、意味的に間接的な関係にあると考えられる節。修飾節と被修飾名詞の間に何らかの説明が省略されていると考えられるタイプ。」と説明している。

《(その他)》－(内容節)》

- (7) 一九九七年に「六つの目しかない骰子を振りながら、七つ目を求めているような表現者でありたい」との願いを込めて発足したグループ。[PN4g_00002, file00029]

事例(7)は前項の事例(1)で挙げた「節＋の＋名詞」と同じ解釈であるが、名詞修飾節の下位分類では、もう 1 点、別の問題がある。「鳥バンク」では、2 段階で「節間キーワード」を「節＋の＋名詞」とする「その他」とは別に「用言＋接続表現＋の」という分類を設けてお

り、「用言＋接続表現＋の」という形式で名詞を修飾する節」と説明している。日本語の文例には「これは次回へ備えての休息期間である。」などをあげている。この2種類の分類項目の区別は極めて不明瞭であり、従属節の意味分類に混乱をもたらすものであるが、「鳥バンク」の修正により解消できる。

《(内容節)－(縮約形修飾節)》

- (8) 不要なごみを減らす社会づくりに貢献する中小企業への対策が急がれる。[PN4g_00002, file00030]

事例(8)は「不要なごみを減らす」が名詞修飾節であり「社会」が被修飾名詞であるが、内容節とするか縮約形修飾節とするか判断に迷うところである。益岡・田窪 (1992) では、「鳥バンク」で縮約形修飾節とされているものは内容節に含まれている。名詞修飾節を詳細に分類することの要否については議論の余地がある。

B. 副詞節における2段階以下の齟齬

副詞節の2段階以下で生じた齟齬の42例中、手段－因果関係(原因)21例が最多で、他に因果関係の下位分類(原因－結果)5例、(時－条件)3例などがある。副詞節の下位分類において齟齬が生じたものの中で、作業者の一方が因果関係とタグ付けしたものが27例、条件・譲歩とタグ付けしたものが5例、付帯状況とタグ付けしたものが5例であり、齟齬のバリエーションは多くない。以下に高頻度の齟齬について事例を挙げて検討する。

《(手段)－(因果関係：原因)》

- (9) 信号の約6万カ所を集中制御タイプなどに切り替え、約150万トンのCO₂が減らせるとする。[PN5a_00003, file00043]

このタイプの齟齬が副詞節の下位分類での齟齬で最多の21例であった。「鳥バンク」では、副詞節(手段)について「主節の内容を実現するための手段に従属節で表す。」としている。副詞節(因果関係)については「従属節と主節で表される事態間の因果関係を表す。」とし、さらにその下位分類である(因果関係：原因)については「従属節の内容が原因となって主節の内容が起こること示す。」と説明している。この分類には非常に悩ましいところがあり、分類の要否について議論の余地がある。

《(因果関係：結果)－(因果関係：原因)》

- (10) あちこちで芝がはがれピッチ上に穴があくような状況は見られなかった。[PN2c_00002, file00020]

「鳥バンク」の因果関係の下位分類である(結果)と(原因)については、(結果)を「従属節の内容が起こった結果、主節の内容が起こることを表す。」とし、(原因)を「従属節の内容が原因となって主節の内容が起こること示す。」としている。結果と原因のどちらを重視しているかという書き手(話し手)の意図を汲取って分類する必要がある、この分類の要否についてもまた議論の余地がある。

《(時) — (条件・譲歩)》

(11) 青空の下で汗ばみながら、砂利を掘って設置し終わると、スピーカーで鳴き声を流す。
[PN5b_00002, file00031]

「鳥バンク」では、従属節の節末形式「〜と」を伴う副詞節を2段階で「時」と「条件」に分類している。これは英語の *when, if* の意味に関連すると思われる。節の説明では、時は「従属節と主節が表す事態の時間的関係を表す。」として、条件・譲歩は「従属節と主節で表される事態間の依存関係を表す。」としている。時の例文として「ベルがなるとすぐに芝居が始まった。」を挙げ、条件・譲歩の例文として「人を助けると気持ちが良い」を挙げている。しかし、日本語学的には副詞節の節末形式「と」は条件節に分類され、益岡・田窪 (1992) では、条件・譲歩を表す副詞節の説明の中で、「花子は、家に帰ると、すぐに友人に電話をかけた。」という例文を挙げて、「すでに成立した個別的事態についての依存関係を表すこともできる。」(p.193) としている。

2. 4 日本語を対象とした場合の「鳥バンク」の意味分類上の問題点

「鳥バンク」節間意味分類体系に則った意味分類アノテーションのタグ付けにおいて生じた齟齬に着目して分析することより、以下のような問題点が明らかになった。

第一に、全ての従属節において2段階以下に機能的分類、意味的分類、形式的分類の混在が見られる点である。1例を挙げると、名詞修飾節では2段階で機能と形式の混在、4段階では機能、意味、形式の混在が見られるが、名詞修飾節の下位で齟齬が認められた事例 (7) のような場合には、意味と形式を対立させる齟齬が生じた。

第二に、意味の分類基準である。「鳥バンク」は英語-日本語翻訳を意識して体系化されたものであるため、意味分類項目も英語翻訳に適用しやすいものとなっている。事例 (11) はその例であり、日本語の意味分類の基準とするためには、日本語学の視点での再構築を必要とする。

第三に、「節間キーワード」の問題である。これは上述した意味分類基準とも関わるが、複数の従属節に同形の「節間キーワード」があり、同義と思われるものも多いため、意味分類作業での齟齬の要因になったと考えられる。また、「節間キーワード」は主に節末形式であるが、事例 (7) のように節境界を跨いでいる場合は修正の必要がある。

3. 「鳥バンク」節間意味分類体系の修正の検討

意味分類タグ付け作業で生じた齟齬と「鳥バンク」意味分類体系との関連についての議論を踏まえ、「鳥バンク」節間意味分類体系の階層の見直し、分類項目・「節間キーワード」の整理・修正などについて検討を行う。

3. 1 意味分類階層についての修正

「鳥バンク」では、従属節の意味・機能を4段階に分類しているが、2段階以下に意味、機能、形式の混在が見られる。これを1, 2, 3段階は機能、4, 5段階は意味・形式での分類となるよう整理し、形式での分類は5段階で統一して行う。また、「節間キーワード」は意味分類の手掛かりとして非常に有用であるが、「鳥バンク」の「節間キーワード」は節末表現と、節外の語（節直後や主文末）が混在しているため、節末に限定して整理する。（節外のものは注記する。）分類についての修正提案の主なもの（簡易版）を表4に示す。（着色

したセルはその段階での分類を行わないものである。)

表 4 意味分類の階層

機能的分類			意味的・形式的分類		節末キーワード
1 段階	2 段階	3 段階	4 段階	5 段階	
補足節	補足節			コト型	コト＋格助詞 コト＋とりたて助詞
				ノ型	ノ＋格助詞 ノ＋とりたて助詞
				トコロ型	トコロ＋格助詞 トコロ＋とりたて助詞
				節＋格助詞型	格助詞
	疑問節	選択疑問文			か／や／やら／かしら
		疑問語疑問文			疑問語＋か
	引用節	直接引用			」と
					」など（と）
					」なんて
		間接引用		ト型	と
				ヨウ型	よう（に）
				ナド型	など（と）
				ナンテ型	なんて
連体節	補足語修飾節	限定／非限定			
	内容節	同格関係			（という）
				節＋の	の
		非同格関係			（ような）
				節＋の	の
	形式名詞修飾節				（節外）ため・もの・はず・・・
連用節	副詞節		時間節		とき・ころ・際・・・
			条件・譲歩		と・ば・たら・・・
			原因・理由		ので・ため・から・・・
			付帯状況		ながら・つつ・・・
			様態		ように・ふうに・・・
			逆接		が・けれど・・・
			手段		て
			目的		ために・ようと・・・
			その他		
	並列節		順接的並列		連用中止・て・・・
			逆接的並列		が・けれど(も)・・・

3. 2 補足節に関する修正提案

補足節の下位分類である名詞節を補足節に改める。これは格助詞までを節として捉えるためである。「鳥バンク」では補足節の下位の名詞節で格助詞の付かないもの（日本語の文例「来年からそれは廃止になるとの事だ」（下線は筆者による）など）も補足節に分類しているが、これは補足節としない。「鳥バンク」の名詞節の説明に基づいて検討すると、「廃止になるとの」が形式名詞「事」の名詞修飾節であることでは、「節＋形式名詞」の形で名詞相当の働きを持つ」という補足節（名詞節）の説明に合致している。しかし「格助詞を伴って主節の述部を補う要素となる節」という説明には合致していないため、「ことだ」「のだ」などについては補足節に分類しないとの修正を行う。

引用節の下位には、「と」「よう」「など」「なんて」を伴った引用形式での分類があれば、「鳥バンク」の選択疑問節型、疑問語疑問節型という分類は不要であると考える。

3. 3 名詞修飾節（連体節）に関する修正提案

1 段階の名称を連体節とする。これは、後述する連用節の提案に伴うもので、連体節と連用節に分類した方が対比的で分かりやすいためである。

補足語修飾節の下位での限定・非限定の分類については、意味が文脈に依存するため、分類の要否には議論の余地がある。

形式名詞を修飾する場合は、分類基準も明確であるため、形式名詞修飾節を独立させた。被修飾語となる形式名詞は「節間キーワード」に（節外）と注記して列挙しておく。

上記の名詞修飾節以外は 2 段階では全て内容節に分類し、その下位において同格か否か、「の」で接続されるか否かで分類する。

3. 4 副詞節・並列節に関する修正提案

副詞節・並列節に関する最大の修正提案は、副詞節・並列節を連用節の下位に分類することである。丸山 (2016) では、日本語文法の先行研究における従属節の分類を比較し、「英語などの言語では、副詞節と並列節（等位節）は形態的に大きく異なるため、両者を区分する必然性があるが、日本語の並列節の場合、（中略）形態的には連用接続形式の一種とみなしてよい。」(p.1114) としている。本研究では、丸山(2016)に倣い、従属節の 1 段階では補足節・連体節・連用節に 3 分類する。

連用節・並列節においては、機能の分類は 2 段階までとし、4 段階で意味の分類を行う。意味・機能をどの深さまで分類するかという議論については今後の課題とする。

3. 5 節間キーワードに関する修正提案

「節間キーワード」については 2 種類の重複の修正を提案する。1 つは複数の節に同形の「節間キーワード」が表記されている問題の修正であり、もう 1 つは一つの「節間キーワード」の中に別の節の「節間キーワード」が含まれている問題の修正である。

以下の「鳥バンク」の日本語の文例のように、意味・機能の異なる節が同形の節末形式で表現されることは珍しくないため、混乱を招かないための表記の工夫が必要である。

なお、「鳥バンク」の日本語の文例 (12)(13)(15)の「節間キーワード」には波線を付した。

u

(12) 私は友達の顔色の悪いのに気付いた。《補足節（名詞節：ノ型）》

(13) 1 時には来るはずなのに友達はまだ来ない。《副詞節（逆接）》

(12)(13)の日本語の文例では、いずれも「節間キーワード」が「のに」になっており、表層的には節末形式は同形である⁹。このような場合は、同形の「節間キーワード」が他に存在することを注記するなど、表記の方法を工夫する必要があると考える。実際のアノテーション作業において、同形の「節間キーワード」を持つ複数の従属節があることで意味分類に齟齬が生じた例を挙げる。

《副詞節（条件・譲歩：偶有的）－補足節（名詞節：トコロ型）》

(14) 7月の交通量を比較調査したところ、12時間の通行車両は約 7600 台から約 9100 台に増えたが、[PN5a_00003, file00043]

事例(14)は1次作業で生じた齟齬である。「鳥バンク」では、「ところ」を補足節（名詞節：トコロ型）、副詞節（時：事態の時：進行）、副詞節（条件・譲歩：偶有的）の3ヶ所で扱っている。事例(14)の従属節を補足節とするのは誤分類であるとしても、副詞節の下位で（時：事態の時：進行）か（条件・譲歩：偶有的）かで分類の揺れが生じる可能性は大きい。

前述の事例(2)でも「節間キーワード」の複数節での重複の問題は指摘したが、「節間キーワード」の重複があることが認識できるように表記しておくことが求められる。

次に1つの「節間キーワード」に異なる従属節の「節間キーワード」が含まれている問題について「鳥バンク」の日本語の文例を挙げて説明する。

(15) 一生懸命勉強しただけのことはあって、今度の試験はよくできた。《副詞節（因果関係：原因）》

(15)に挙げた「鳥バンク」の日本語の文例は、「節間キーワード」を下線部「だけのことはあって」としているが、この中には「だけ」という副詞節（程度）の「節間キーワード」と「節十の十名詞」という名詞修飾節の「節間キーワード」が含まれていると見ることができる。文脈上、節末の「て」だけでも十分に同様の《副詞節（因果関係：原因）》意味分類が可能である。個別の節末形式を、基準を設けずに「節間キーワード」に加えていくと、節認定の問題も含み、分類が複雑になり過ぎる危険がある。

4. おわりに

本発表では、節境界アノテーション・従属節の意味分類タグ付けに関する作業の問題点を整理した松本・浅原・有田 (2017) で得たデータに基づき、意味分類タグの付与に齟齬があった事例を再検討し、タグ付け作業の基準とした「鳥バンク」の再構築について議論した。

1段階の従属節の機能分類における齟齬では、9割が副詞節－並列節で生じており、その他の補足節－名詞修飾節、補足節－副詞節で生じた齟齬も引用節を含む特定のものであった。2段階以下の分類の齟齬では副詞節の下位での齟齬が最も多く、その5割が手段－因果関係で生じ、時－条件や因果関係の下位分類である原因－結果においても齟齬があった。これらの齟齬を精査することで英語への翻訳を意識した「鳥バンク」意味分類体系を日本語に適用したことによるいくつかの問題が明らかになった。そこで「鳥バンク」を日本語仕様に

⁹ 節認定の問題が含まれるが、節認定については別稿に譲る。

再構築するための修正について具体的事例を挙げて示した。

修正提案としては、まずは機能と意味分類の階層である。1段階においては、「鳥バンク」の補足節（補語相当節）・名詞修飾節（連体節）・副詞節（副詞的連用節）・並列節（並列型連用節）の4分類を、補足節・連体節・連用節の3分類に修正する。これにより、1段階での従属節の機能分類における問題は解決できる。

従属節の意味や節末形式に依拠した分類については、「鳥バンク」の詳細な分類項目や「節間キーワード」を日本語学の視点から見直し、修正を加えることで、日本語仕様の安定した従属節の意味分類基準に再構築できることが明らかになった。

今後は、意味分類で齟齬のなかった節についても分析し、分類の深さや節認定についての議論も行う必要がある。従属節の意味分類についての揺れや齟齬を許容していくという視点で、「鳥バンク」をベースとした漏れのない日本語従属節意味分類基準を策定することを今後の課題としたい。

謝 辞

本研究（の一部）は国立国語研究所コーパス開発センターの共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」（2016-2021 年度）の成果である。

文 献

- 池原悟 (2007) 「鳥バンク (Tori-Bank)」 unicom.ike.tottori-u.ac.jp/toribank
- 池原悟 (2009) 『非線形言語モデルによる自然言語処理』 岩波書店
- 益岡隆志 (1997) 『複文』 くろしお出版
- 益岡隆志・田窪行則 (1992) 『基礎日本語文法』 くろしお出版
- 松本理美・浅原正幸・有田節子 (2017) 「『現代日本語書き言葉均衡コーパス』に対する節の意味分類情報アノテーションー基準策定、仕様書作成の必要性についてー」 『言語資源活用ワークショップ 2016 発表論文集』 pp.336-346
- 丸山岳彦・佐藤理史・夏目和子 (2016) 「現代日本語における節の分類体系について」 『言語処理学会第 22 回年次大会発表論文集』 pp.1113-1116